Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 777–806
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Endogenous selection or treatment model estimation ☆

Arthur Lewbel*

Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467, USA

Available online 17 January 2007

Abstract

In a sample selection or treatment effects model, common unobservables may affect both the outcome and the probability of selection in unknown ways. This paper shows that the distribution function of potential outcomes, conditional on covariates, can be identified given an observed variable V that affects the treatment or selection probability in certain ways and is conditionally independent of the error terms in a model of potential outcomes. Selection model estimators based on this identification are provided, which take the form of simple weighted averages, GMM, or two stage least squares. These estimators permit endogenous and mismeasured regressors. Empirical applications are provided to estimation of a firm investment model and a schooling effects on wages model.

JEL classification: C14; C25; C13

Keywords: Sample selection; Treatment effects; Censoring; Semiparametric; Endogeneity; Instrumental variables; Switching regressions; Heteroscedasticity; Latent variable models; Ordered choice; Investment; Returns to schooling

☆Portions of this paper were previously circulated under other titles including, "Two Stage Least Squares Estimation of Endogenous Sample Selection Models."

*Tel.: +1 617 552 3678; fax: +1 617 552 2308.
E-mail address: lewbel@bc.edu.
URL: http://www2.bc.edu/~lewbel/.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.11.004

1. Introduction

Assume that for a sample of individuals $i = 1, \ldots, n$ we observe an indicator D that equals one if an individual is treated, selected, or completely observed, and zero otherwise. If $D = 1$ we observe some scalar or vector P, otherwise let $P = 0$. Define P^* to equal the observed P when $D = 1$, otherwise P^* equals the value of P that would have been observed if D had equaled one, that is, either a counterfactual or an unobserved response. Then $P = P^*D$. We also observe covariates, though selection on observables is not assumed. Treatment or selection D may be unconditionally or conditionally correlated with P^*, so P^* and D may depend on common unobservables. Rubin (1974) type restrictions like unconfoundedness or ignorability of selection are not assumed.

To illustrate, in a classic wage model (Gronau, 1974; Heckman, 1974, 1976a,b), $D = 1$ if the individual is employed, P^* is the wage an individual would get if employed, and P is the observed wage, which is zero for the unemployed. Both P^* and D depend on common unobservables such as ability, as well as on observable covariates such as measures of schooling or training.

Another example is models based on data sets where some regressors are missing, not at random. For example, models of individual's consumption or purchasing decisions depend on income P^*, and in surveys many individuals do not report their income. Failure to report income ($D = 0$) is likely to be correlated with income, even after conditioning on other observed covariates.

For simplicity, refer to D as selection, though more generally it is just an indicator of not observing P^* for whatever reason. Assume that selection D is given by

$$D = I(0 \leqslant M^* + V \leqslant A^*), \tag{1}$$

where the unobserved A^* can be a constant, a random variable, or infinity, V is an observed, continuously distributed covariate (or known function of covariates) with large support, M^* is an unobserved latent variable, and I is the indicator function that equals one if its argument is true and zero otherwise. Typical parametric or semiparametric models of selection are special cases of Eq. (1) where M^* is linear in covariates X and a well behaved error term e, but that structure is not imposed here.

Setting the lower bound to zero in Eq. (1) is a free normalization, since no location assumptions are imposed on M^* and A^*. Similarly, setting the coefficient of V to one is (apart from sign) a free scale normalization that is imposed without loss of generality.

In the wage model example, the typical assumption is that one chooses to work if the gains in utility from working, indexed by the latent $M^* + V$, are sufficiently large, so in that case A^* is infinite. Examples in which A^* is finite arise in ordered treatment or ordered selection models. For example, if an ordered choice model with latent variable $M^* + V$ determines an individual's years of schooling and D indexes having exactly 12 years of schooling then individuals with $M^* + V < 0$ choose 11 or fewer years while those with $M^* + V > A^*$ choose 13 or more years. We might then be interested in modeling the returns P from having exactly 12 years of schooling. Examples of models like this with A^* random include Cameron and Heckman (1998) and Carneiro et al. (2003).

A convenient feature of the proposed estimators is that they will not require specifying, modeling or estimating the D (propensity score) model, apart from assuming Eq. (1) holds. For example, any dependence of M^* on X can be unknown and need not be estimated, and